Preprints.org

Sparse Boosting for Additive Spatial Autoregressive Model with High Dimensionality

[Mu Yue](#) * and [Jingxin Xi](#)

Article

Sparse Boosting for Additive Spatial Autoregressive Model with High Dimensionality

Mu Yue [1,*] and Jingxin Xi [1,2]

[1] School of Physical and Mathematical Sciences, Nanyang Technological University
[2] School of Ecology & Environment, Renmin University of China
[*] Correspondence: yuemu.moon@gmail.com

Abstract: Variable selection methods have been a focus in the context of econometrics and statistics literature. In this paper, we consider additive spatial autoregressive model with high-dimensional covariates. Instead of adopting the traditional regularization approaches, we offer a novel multi-step sparse boosting algorithm to conduct model-based prediction and variable selection. One main advantage of this new method is that we do not need to perform the time-consuming selection of tuning parameters. Extensive numerical examples illustrate the advantage of the proposed methodology. An application of Boston housing price data is further provided to demonstrate the proposed methodology.

Keywords: sparse boosting; variable selection; spatial autoregressive model; additive model; instrument variable

---

1. Introduction

Spatial data, also known as geospatial data, commonly appear in fields such as environmental health, economics, and epidemiology. This type of data can be represented by numerical values in a geographic coordinate system. The spatial autoregressive (SAR) model, which includes a spatially lagged term of the response variable to account for spatial dependence among spatial units, has become an important research topic in recent econometrics and statistics literature. Significant developments in the estimation and inference of this model can be found in the works of Cliff and Ord [1], Anselin [2], Robert [3], among others.

The linear SAR model, which extends the ordinary linear regression model by incorporating a spatially lagged term of the response variable, has been extensively studied [4–8]. However, it imposes assumptions that might be unrealistic in practice, rendering it inefficient if not correctly specified. To enhance model flexibility and adaptability, the nonparametric SAR model allows for investigating the relationship between the response variable and predictors without assuming a specific shape for the relationship. This model has recently gained much attention from econometricians and statisticians. One of the most powerful and useful models in spatial statistics is the partial linear SAR model, where the spatially lagged response variable and some predictors enter the model linearly, while the remaining predictors are incorporated nonparametrically. Specifically, Su and Jin [9] proposed a profile quasi-maximum likelihood estimation method for the model. Koch and Krisztin [10] proposed an estimation method based on B-splines and genetic algorithms. Chen et al. [11] proposed a two-step Bayesian approach based on kernel estimation and Bayesian methods for inference. Krisztin [12] proposed a novel Bayesian semiparametric estimation method combining penalized splines with Bayesian methods. Li and Mei [13] proposed a method to test linear constraints on the parameters of the partially linear SAR model.

Beyond the partial linear SAR model, Su [14] studied the SAR model, where the spatially lagged response variable enters the model linearly and all predictors enter nonparametrically, proposing semiparametric GMM estimation under weak moment conditions. Wei et al. [15] studied SAR models with varying coefficients to capture heterogeneous effects of covariates and spatial interaction. Du et al. [16] considered a class of partially linear additive SAR models and proposed a generalized

method of moments estimator. Cheng and Chen [17] studied partially linear single-index SAR models and proposed profile maximum likelihood estimators. In this paper, we will consider the additive SAR model, which is more flexible than the SAR model examined in [14], as it includes only one nonparametric component, easing the curse of dimensionality issue. In particular, we will address high-dimensional data settings where the dimensionality of predictors may exceed the sample size.

Data science is an ever-expanding field. High-dimensional models, where the feature dimension grows exponentially or non-polynomially fast with sample size, have become a focus in statistical literature. Including irrelevant features in the model can lead to undesirable computational issues and unstable estimation. To address these challenges, variable screening and selection techniques have been developed. Among these developments, penalized approaches such as lasso [18], SCAD [19], MCP [20], and their various extensions [21–23] have been thoroughly studied to identify important covariates and estimate the coefficients of interest simultaneously, thereby improving the prediction accuracy and interpretability of statistical models. For example, following the idea of group lasso [23] and adaptive lasso [22], Wei et al. [15] proposed a local linear shrinkage estimator for the SAR model with varying coefficients for model selection. Nevertheless, these regularization methods have one major disadvantage: they all involve tuning parameters that must be chosen using computationally intensive methods such as cross-validation.

In recent decades, boosting methods have become an effective alternative tool for high-dimensional data settings to perform variable selection and model estimation. They offer several advantages, such as relatively smaller computational cost, lower risk of overfitting, and simpler adjustments to incorporate additional constraints. Boosting was initially conceptualized as a machine learning algorithm that constructs better base learners to minimize the loss function with every iteration. The original version of the boosting algorithm proposed by Schapire [24] and Freund [25] did not fully exploit the potential of base learners. This changed when Freund and Schapire [26] proposed the AdaBoost algorithm, which could adapt to base learners more effectively. Following the development of this algorithm, many other boosting algorithms have been formulated, with major variations in their loss functions. For instance, AdaBoost with the exponential loss, L2Boosting [27] with the squared error loss, SparseL2Boosting [28] with the penalized loss, and HingeBoost [29] with the weighted hinge loss. Other versions of boosting algorithms proposed recently include pAUCBoost [30], Twin Boosting [31], Twin HingeBoost [29], ER-Boost [32], GSBoosting [33], among others. As demonstrated by Yue et al. [34–37], sparse boosting achieves better variable selection performance than L2Boosting as well as some penalized methods such as lasso. However, the application of (sparse) boosting approaches in high-dimensional SAR models has not yet been investigated.

However, there currently exists a lack of exploration regarding the application of (sparse) boosting methods in the study of high-dimensional spatial autoregressive (SAR) models, which provides an opportunity for further investigation in this research. This paper focuses on additive spatial autoregressive models with high-dimensional covariates and proposes a novel multi-step sparse boosting algorithm aimed at model-based prediction and variable selection, rather than relying on traditional regularization methods. A significant advantage of this new method is that it eliminates the need for time-consuming tuning parameter selection, thereby enhancing the convenience of modeling. Our approach is designed specifically for high-dimensional additive spatial autoregressive models, improving the flexibility and adaptability of the model in complex situations. Extending existing boosting techniques to such complex model structures presents a considerable technical challenge and requires deeper exploration both theoretically and practically. The detailed research presented in this paper facilitates the first application of this method in the field, thereby providing a foundation for future research. Through simulation studies and real data examples, we demonstrate the superior performance of the proposed method compared to various alternative algorithms. Our results not only fill the existing research gap in the literature but also offer empirical support for the multi-step sparse boosting method as an effective tool for high-dimensional data analysis, especially in scenarios involving complex relationships among variables.

The rest of the paper is organized as follows. In Section 2, the additive autoregressive model is formulated, and a multi-step sparse boosting algorithm is proposed. In Section 3, simulation studies are conducted to demonstrate the validity of this multi-step method. In Section 4, the performance of the multi-step sparse boosting method is evaluated by analyzing Boston housing price data. Concluding remarks are given in Section 5.

2. Methodology

2.1. Model and Estimation

Consider the following additive spatial autoregressive model:

$$Y = \rho WY + \sum_{j=1}^{p} g_j(X_j) + \epsilon, \tag{1}$$

where $Y = (y_1, \cdots, y_n)^T$ is the $n \times 1$ vector of the observations of the response variable with n being the number of spatial units, W is an $n \times n$ spatial weight matrix of known constants with zero diagonal elements and WY is referred to as spatial lag of Y, ρ is the spatial autoregressive parameter with $|\rho| < 1$, $X_j = (X_{1j}, \cdots, X_{nj})^T$ is the $n \times 1$ vector of the j-th regressor and $X = (X_1, \cdots, X_p)$ is the $n \times p$ observed matrix of regressors. For simplicity, the covariates $X_1, \cdots, X_p$ are assumed to be distributed on compact intervals $[a_1, b_1], \cdots, [a_p, b_p]$ respectively. $g_j(.)s$ are unknown smooth functions on $[a_j, b_j]$ with the assumption that $E[g_j(X_j)] = 0$ for identifiability purpose. $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T$ is an $n \times 1$ vector of i.i.d disturbances with zero mean and finite variance σ^2.

The smoothing splines technique is typically used to estimate the unknown functions, demonstrating steady performance in practice. In this paper, we will use B-spline basis functions to approximate the unknown coefficient functions $g_j(.), j = 1, \cdots, p$. Let $B(.) = (B_1(.), \ldots, B_L(.))$ be an equally spaced B-spline basis, where L is the dimension of the basis. Similar to Huang et al. [38], let $\bar{B}_{lj} = \frac{1}{n} \sum_{i=1}^{n} B_l(X_{nj})$, $\pi_{lj}(.) = B_l(.) - \bar{B}_{lj}$, $\pi_j(.) = (\pi_{1j}(.), \cdots, \pi_{Lj}(.))$. Then under appropriate smoothness assumptions, $g_j(.) \approx \sum_{l=1}^{L} \pi_{lj}(.)\alpha_{lj} = \pi_j(.)\alpha_j$ for $j = 1, \cdots, p$. Then the spline estimator of $g(.) = \sum_{j=1}^{p} g_j(.)$ is $\hat{g}(.) = \pi(.)\alpha$, where $\pi(.) = (\pi_1(.), \cdots, \pi_p(.))$, $\alpha = (\alpha_1^T, \cdots, \alpha_p^T)^T$. It is obvious that $\sum_{i=1}^{n} \hat{g}_j(X_{ij}) = 0$ for $j = 1, \cdots, p$.

Then the model (1) can be rewritten as follows

$$\begin{aligned} Y &\approx \rho WY + \sum_{j=1}^{p} \pi_j(X_j)\alpha_j + \epsilon \\ &= \rho WY + \pi\alpha + \epsilon, \end{aligned} \tag{2}$$

where $\pi = (\pi_1(X_1), \cdots, \pi_p(X_d))$. Let $P = \pi(\pi^T \pi)^{-1}\pi^T$ denote the projection matrix onto the space spanned by π. Similar to [39], partialing out the B-spline approximation, we obtain

$$(I - P)Y \approx (I - P)\rho WY + (I - P)\epsilon. \tag{3}$$

In the above equation, a problem of endogeneity emerges because the spatially lagged value of Y is correlated with the stochastic disturbance. Suppose H is a relevant vector instrument to eliminate the endogeneity of WY. Let $D = WY$. Regressing D on instrument H via OLS produces the fitted value for the endogenous variable D:

$$\hat{D} = H(H^T H)^{-1} H^T D. \tag{4}$$

Substituting D with its predicted value in model (3), it becomes

$$\begin{aligned} (I - P)Y &\approx (I - P)\rho H(H^T H)^{-1} H^T D + (I - P)\epsilon \\ &= (I - P)\rho MD + (I - P)\epsilon, \end{aligned} \tag{5}$$

where $M = H(H^T H)^{-1} H^T$. Then by OLS,

$$\hat{\rho} = (D^T M (I - P) M D)^{-1} D^T M (I - P) Y. \tag{6}$$

Substituting D with $\hat{D}$ and ρ with $\hat{\rho}$ in model (2), it becomes

$$Y \approx \hat{\rho} D + \pi \alpha + \epsilon. \tag{7}$$

Then the least squares loss function is close to

$$(Y - \hat{\rho} D - \pi \alpha)^T (Y - \hat{\rho} D - \pi \alpha). \tag{8}$$

However, when dimensionality of π is larger than sample size n, least square estimation fails. In this case, we will adopt sparse boosting approach to estimate α. Denote $\hat{\alpha}$ as the estimator of α obtained through sparse boosting, using the squared loss function (8) as the loss function. The detailed sparse boosting algorithm will be given in the next subsection.

Consequently, $g_j(.)$ can be estimated by $\hat{g}_j(.) = \pi_j(.)\hat{\alpha}_j$ for $j = 1, \cdots, p$. The variance σ^2 can be estimated by

$$\hat{\sigma}^2 = \frac{1}{n}(Y - \hat{\rho}WY - \pi\hat{\alpha})^T(Y - \hat{\rho}WY - \pi\hat{\alpha}). \tag{9}$$

Similar to [39], we use an analogous rationality for the construction of instrument variables. In the first step, we simply regress Y on pseudo regressors variables WY and π. Then the least squares loss function is

$$(Y - \rho WY - \pi\alpha)^T(Y - \rho WY - \pi\alpha) \tag{10}$$

Denote $(\tilde{\rho}, \tilde{\alpha})$ as the estimators of (ρ, α) obtained through sparse boosting, using the squared loss function (10) as the loss function. Then, we can use the following instrumental variable:

$$\tilde{H} = W(I - \tilde{\rho}W)^{-1}\pi\tilde{\alpha}. \tag{11}$$

In the second step, instrumental variable $\tilde{H}$ is used to obtain the estimators $\bar{\alpha}$ and $\bar{\rho}$, which are then used to construct the instrumental variable

$$H = W(I - \bar{\rho}W)^{-1}\pi\bar{\alpha}. \tag{12}$$

Finally, use the instrumental variables H to obtain the final estimators $\hat{\rho}$ and $\hat{\alpha} = (\hat{\alpha}_1^T, \cdots, \hat{\alpha}_p^T)^T$. The function $g_j(.)$ can be estimated by $\hat{g}_j(.) = \hat{\pi}_j(.)\hat{\alpha}_j$ for $j = 1, \cdots, p$ and the response Y can be estimated by $\hat{Y} = (I - \hat{\rho}W)^{-1}\pi\hat{\alpha}$.

2.2. Sparse Boosting Techniques

Sparse boosting can be viewed as iteratively pursuing gradient descending in function space using a penalized empirical risk function that integrates squared loss and the complexity of the boosting measure. Similar to Yue et al. ([34–37]), we will adopt the g-prior minimum description length (gMDL) [40], a combination of squared loss and the trace of boosting operator, as the penalized empirical risk function to estimate the update criterion in each iteration and the stopping criterion. We use it because it has a data driven penalty to avoid the selection of the tuning parameter. To facilitate presentation, suppose the vector Y^* is regressed on the p^*-dimensional matrix $X^* = (X_1^*, \cdots, X_p^*)$.

Then the least squares loss function involved in sparse boosting is $(Y^* - X^*\theta^*)^T(Y^* - X^*\theta^*)$, where $\theta^* = ((\theta_1^*)^T, \cdots, (\theta_p^*)^T)^T$. The gMDL takes the form:

$$\text{gMDL}(Y^*, RSS, \text{trace}(\mathcal{B})) = \log(F) + \frac{\text{trace}(\mathcal{B})}{n}\log(\frac{Y^{*T}Y^* - RSS}{\text{trace}(\mathcal{B}) \times F}),$$

$$F = \frac{RSS}{n - \text{trace}(\mathcal{B})}, \tag{13}$$

where RSS is the residual sum of squares and $\mathcal{B}$ is the boosting operator. The model achieve shortest description of data will be chosen.

We present the sparse boosting approach more specifically. The initial value of θ^* is set to the zero vector, i.e. $\theta^{*[0]} = 0$. In each of the kth iteration ($0 < k \le K$, and K is the maximum number of iterations considered in the first step), we use the residual $R^{[k]} = Y^* - X^*\theta^{*[k-1]}$ from the current iteration to fit each of the j-th component $X_j^*, j = 1, \cdots, p^*$. The fit, denoted by $\hat{\lambda}_j^{[k]}$, is calculated by minimizing the squared loss function $(R^{[k]} - X_j^*\lambda)^T(R^{[k]} - X_j^*\lambda)$ with respect to λ. Therefore, the least squares estimate is $\hat{\lambda}_j^{[k]} = [(X_j^*)^T(X_j^*)]^{-1}(X_j^*)^T R^{[k]}$, the corresponding hat matrix is $\mathcal{H}_d = (X_j^*)[(X_j^*)^T(X_j^*)]^{-1}(X_j^*)^T$ and the residual sum of squares is $RSS_j^{[k]} = (R^{[k]} - X_j^*\hat{\lambda}_j^{[k]})^T(R^{[k]} - X_j^*\hat{\lambda}_j^{[k]})$. The chosen element $\hat{s}_k$ is attained by:

$$\hat{s}_k = \text{argmin}_{1 \le j \le p^*}\text{gMDL}(Y^*, RSS_j^{[k]}, \text{trace}(\mathcal{B}_j^{[k]})), \tag{14}$$

where $\mathcal{B}_j^{[1]} = \mathcal{H}_j$ and $\mathcal{B}_j^{[k]} = I - (I - \mathcal{H}_j)(I - \nu\mathcal{H}_{\hat{s}_{k-1}})\cdots.(I - \nu\mathcal{H}_{\hat{s}_1})$ for $k > 1$ is the first step boosting operator for choosing jth element in the kth iteration. Hence, there is an unique element $X_{\hat{s}_k}^*$ to be selected at each iteration, and only the corresponding coefficient vector $\theta_{\hat{s}_k}^{*[k]}$ changes, i.e., $\theta_{\hat{s}_k}^{*[k]} = \theta_{\hat{s}_k}^{*[k-1]} + \nu\hat{\lambda}_{\hat{s}_k}^{[k]}$, where ν is the pre-specified step-size parameter. All the other $\theta_j^{*[k]}$ for $j \neq \hat{s}_k$ keep unchanged. We repeat this procedure for K times and the number of iterations K can be estimated by

$$\hat{K} = \text{argmin}_{1 \le k \le K}\text{gMDL}(Y^*, RSS_{\hat{s}_k}^{[k]}, \text{trace}(\mathcal{B}^{[k]})), \tag{15}$$

where $\mathcal{B}^{[k]} = I - (I - \nu\mathcal{H}_{\hat{s}_k})\cdots.(I - \nu\mathcal{H}_{\hat{s}_1})$.

From the sparse boosting, we get the estimator of θ^* by $\theta^{*[\hat{K}]} = ((\theta_1^{*[\hat{K}]})^T, \cdots, (\theta_p^{*[\hat{K}]})^T)^T$.

For least squares loss function (8), $Y^* = Y - \hat{\rho}D$ and $X^* = \pi$, and for least squares loss function (10), $Y^* = Y$ and $X^* = (WY, \pi)$.

Overall, as illustrated in Figure 1, the flowchart visually summarizes the methodologies and steps articulated in this paper, thereby enhancing the understanding of our proposed multi-step sparse boosting algorithm for high-dimensional additive spatial autoregressive models.

In Step I, the algorithm starts with parameter initialization, followed by iterations to identify optimal variables using the g-prior minimum description length (gMDL) criterion. Coefficients are updated until a stopping criterion is met, resulting in covariance matrix estimates.

Step II applies the sparse boosting algorithm again to refine variable selection and update coefficients. This approach enhances model accuracy and selection efficiency, making it well-suited for high-dimensional data analysis.
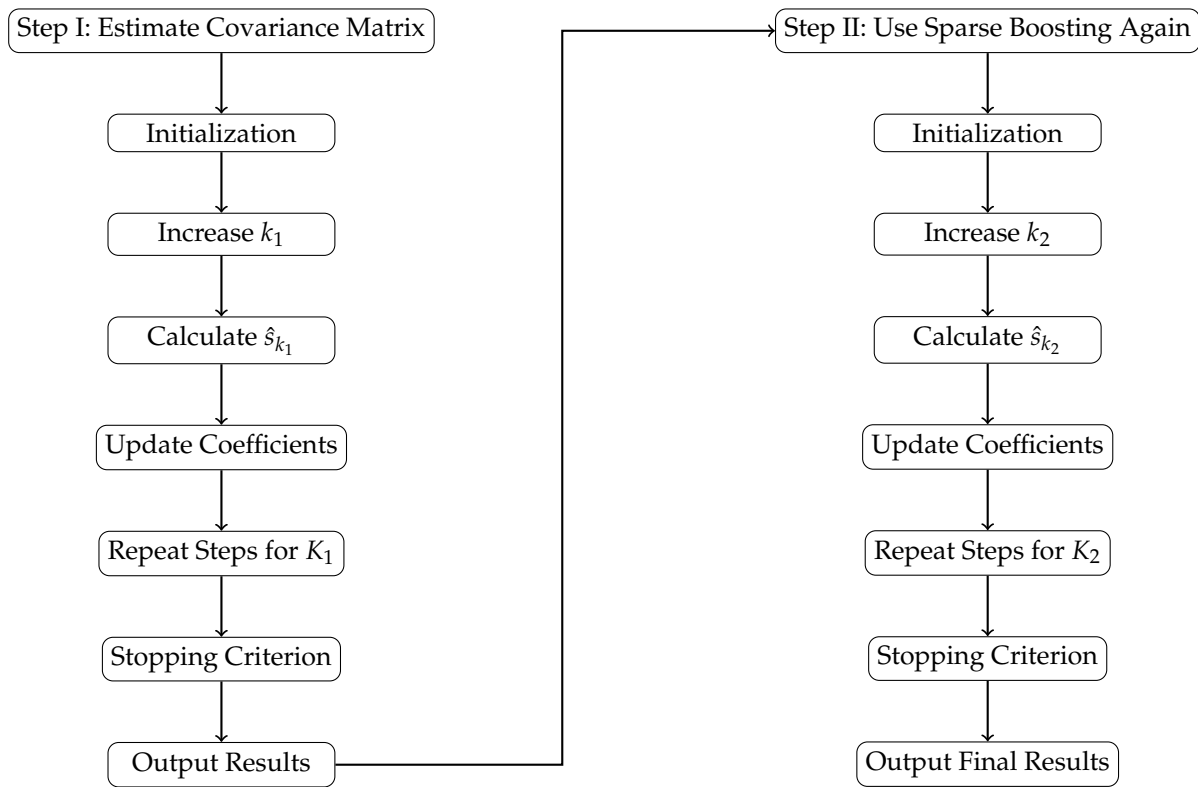
Figure 1. Flowchart of the Sparse Boosting Algorithm: The methodological framework and computational processes of the sparse boosting algorithm, facilitating comprehension of the key steps involved.

3. Simulation

In this section, we investigate the finite sample performance of the proposed methodology with Monte Carlo simulation studies. The data is generated from the following model:

$$Y = \rho WY + \sum_{j=1}^{1000} g_j(X_j) + \epsilon, \tag{16}$$

where $X_1, \cdots, X_{1000}$ are all i.i.d on $[-1, 1]$. $g_1(x) = g_3(x) = 2\tan(x)$, $g_2(x) = g_4(x) = 2\sin(x)$ and $g_5(x) = \cdots g_{1000}(x) = 0$. The error term ϵ follows a normal distribution with mean 0 and variance σ^2. Similar to [41], the weight matrix is set to be $W = I_R \otimes E_m$, where $E_m = (l_m^T l_m - l_m)/(m-1)$, l_m is the m-dimensional unit vector, and $\otimes$ is Kronecker product. Thus, the number of spatial units is $n = R \times m$. For comparison, we consider $(R, m) = (10, 10)$ and $(20, 20)$, corresponding to $n = 100$ and 400 respectively. We evaluate three different values of $\rho = 0.2, 0.5, 0.8$, representing weak to strong spatial dependence of the responses. Additionally, we consider $\sigma^2 = 0.5$ and 1. The degree of freedom for B-splines basis is set to be $\lfloor(N/\lfloor N/log(N)\rfloor)\rfloor$, where $\lfloor x \rfloor$ is the integer part of x.

In face of ultra-high dimensional data, pre-screening can be adopted to reduce the dimensionality to a moderate size. In particular, we adopt marginal screening by l^2-norm [42] to screen out irrelevant covariates. More specifically, we regress Y on each covariate to construct the marginal model $Y = \rho_0 WY + g_{0j}(X_j) + \epsilon_{0j}, j = 1, \cdots, p$. The empirical l^2-norm of an estimated function $\hat{g}_{0j}(X_j) = (\hat{g}_{0j}(X_{1j}), \cdots, \hat{g}_{0j}(X_{nj}))^T$ is defined as:

$$\|\hat{g}_{0j}(X_j)\|^2 = \frac{1}{n}(\hat{g}_{0j}(X_{1j}), \cdots, \hat{g}_{0j}(X_{nj}))(\hat{g}_{0j}(X_{1j}), \cdots, \hat{g}_{0j}(X_{nj}))^T. \tag{17}$$

A greater value of this measure suggest a stronger association between the covariate and the response. Following the recommendation by [42–45], the covariates with the largest $\lfloor \frac{n}{\log(n)} \rfloor$ l^2-norm are selected. Thus, for sample size 100 and 400, the first 21 and 66 variables in the ranked list are selected to conduct downstream analysis, respectively. Thereafter, we proceed to use our proposed multi-step sparse

boosting approach to build up the final parsimonious model. For comparison, besides the proposed method using sparse boosting in each step, we also examine other approaches such as l_2-boosting, lasso regression and elastic net regression in each step.

In our implementation, we set the maximum number of boosting iterations to $K = 500$ and the elastic net mixing parameter to 0.5. Penalized methods were executed using the R package **glmnet**, with tuning parameters selected via 5-fold cross-validation. To assess the performance of our approach, we analyze results from 500 replications, reporting the following metrics:

S: coverage probability that the top $\lfloor \frac{n}{\log(n)} \rfloor$ covariates after screening includes all important covariates;

TP: the median of true positives;

FP: the median of false positives;

Size: the median of model sizes;

ISPE: the average of in-sample prediction errors defined as $(Y - \hat{Y})^T ((Y - \hat{Y})/n)$;

RMISE: the average of root mean integrated squared errors defined as

$$\sqrt{\frac{1}{n} \sum_{j=1}^{4} (g_j(X_j) - \hat{g}_j(X_j))^T (g_j(X_j) - \hat{g}_j(X_j))};$$

Bias(ρ): the mean bias of ρ;

Bias(σ): the mean bias of σ.

The simulation results presented in Tables 1 and 2, based on 500 replications, provide comprehensive insights into the performance of various methods for variable screening and selection under two conditions ($\sigma = 0.5$ and $\sigma = 1.0$).

Table 1. Results of variable screening, variable selection, and estimation for simulation when $\sigma = 0.5$. The values in the parentheses are the robust standard deviations. M1: our proposed multi-step sparse boosting method; M2: except sparse boosting, l_2-boosting are used in each step; M3: except sparse boosting, lasso regression are used in each step; M4: except sparse boosting, elastic net regression are used in each step. S: coverage probability that the top $\lfloor \frac{n}{\log(n)} \rfloor$ covariates after screening includes all important covariates; TP: the median of true positive; FP: the median of false positive; Size: the median of model sizes; ISPE: the average of mean square error; RMISE: the average of root mean integrated squared error; Bias(ρ): the mean bias of ρ; Bias(σ): the mean bias of σ. The values in the parentheses are the robust standard deviations; Simulation based on 500 replicates.

n	ρ	Method	S	TP	FP	Size	ISPE	RMISE	Bias(ρ)	Bias(σ)
100	0.2	M1	0.88 (0.32)	4 (0.45)	4 (2.31)	8 (2.31)	1.123 (1.640)	0.754 (0.538)	-0.025 (0.240)	0.393 (0.538)
		M2	0.88 (0.32)	4 (0.20)	12 (2.19)	16 (2.18)	1.212 (1.578)	0.782 (0.496)	-0.032 (0.216)	0.461 (0.507)
		M3	0.88 (0.32)	4 (0.20)	15 (2.42)	19 (2.49)	1.235 (1.500)	0.802 (0.499)	-0.031 (0.228)	0.486 (0.496)
		M4	0.88 (0.32)	4 (0.16)	16 (1.66)	20 (1.70)	1.336 (1.527)	0.843 (0.495)	-0.029 (0.216)	0.535 (0.486)
	0.5	M1	0.92 (0.28)	4 (0.28)	4 (2.50)	8 (2.43)	1.320 (1.657)	0.694 (0.444)	-0.002 (0.112)	0.331 (0.447)
		M2	0.92 (0.28)	4 (0.04)	12 (2.43)	16 (2.43)	1.495 (1.764)	0.744 (0.442)	0.004 (0.118)	0.420 (0.453)
		M3	0.92 (0.28)	4 (0.06)	15 (2.36)	19 (2.36)	1.567 (1.796)	0.773 (0.449)	0.001 (0.109)	0.451 (0.450)
		M4	0.92 (0.28)	4 (0.05)	16 (1.82)	20 (1.82)	1.609 (1.688)	0.797 (0.428)	0.001 (0.107)	0.486 (0.426)
	0.8	M1	0.99 (0.12)	4 (0.08)	4 (1.94)	8 (1.93)	3.127 (2.182)	0.578 (0.197)	0.005 (0.049)	0.204 (0.191)
		M2	0.99 (0.12)	4 (0)	12 (2.29)	16 (2.29)	3.405 (2.099)	0.623 (0.175)	0.006 (0.039)	0.284 (0.158)
		M3	0.99 (0.12)	4 (0)	14 (2.24)	18 (2.24)	3.443 (2.061)	0.641 (0.165)	0.006 (0.038)	0.310 (0.147)
		M4	0.99 (0.12)	4 (0)	16 (1.69)	20 (1.69)	3.64 (2.096)	0.670 (0.167)	0.006 (0.038)	0.354 (0.151)
400	0.2	M1	1 (0)	4 (0.75)	5 (3.10)	9 (3.38)	1.116 (1.674)	0.630 (0.450)	-0.018 (0.691)	0.402 (0.512)
		M2	1 (0)	4 (0)	33 (4.28)	37 (4.28)	1.023 (1.199)	0.589 (0.068)	-0.038 (0.649)	0.376 (0.294)
		M3	1 (0)	4 (0)	52 (8.40)	56 (8.40)	0.841 (1.199)	0.410 (0.153)	-0.039 (0.635)	0.276 (0.329)
		M4	1 (0)	4 (0)	55 (8.24)	59 (8.24)	0.827 (1.008)	0.443 (0.135)	-0.055 (0.604)	0.299 (0.287)
	0.5	M1	1 (0)	4 (0.62)	5 (2.91)	9 (3.12)	1.370 (1.892)	0.596 (0.376)	-0.009 (0.416)	0.360 (0.433)
		M2	1 (0)	4 (0)	33 (4.19)	37 (4.19)	1.261 (1.510)	0.580 (0.050)	-0.026 (0.308)	0.338 (0.166)
		M3	1 (0)	4 (0)	52 (8.55)	56 (8.55)	0.982 (1.156)	0.389 (0.124)	-0.020 (0.331)	0.237 (0.236)
		M4	1 (0)	4 (0)	55 (8.72)	59 (8.72)	1.068 (1.273)	0.427 (0.122)	-0.022 (0.328)	0.268 (0.219)
	0.8	M1	1 (0)	4 (0)	5 (2.25)	9 (2.25)	2.437 (1.593)	0.534 (0.049)	-0.046 (0.134)	0.284 (0.204)
		M2	1 (0)	4 (0)	33 (4.34)	37 (4.34)	2.720 (1.892)	0.579 (0.046)	-0.021 (0.134)	0.331 (0.200)
		M3	1 (0)	4 (0)	52 (8.62)	56 (8.62)	2.299 (1.820)	0.388 (0.116)	-0.023 (0.157)	0.233 (0.279)
		M4	1 (0)	4 (0)	56 (8.83)	60 (8.83)	2.377 (1.832)	0.417 (0.095)	-0.028 (0.109)	0.245 (0.180)

At $\sigma = 0.5$, M1 consistently achieves high coverage probabilities, reaching 0.99 when $\rho = 0.8$, while maintaining a true positive count of 4 and significantly reducing the false positive rates compared to other methods. For example, M2 has a maximum of 12 false positives when $n = 100$ and $\rho = 0.2$, while M3 and M4 peak at 15 and 16 false positives, respectively. In terms of in-sample prediction error (ISPE), M1 demonstrates robust performance with values ranging from 1.123 to 3.127. M2, M3, and M4 show higher ISPE values, indicating a tendency toward overfitting, with M2 reaching 3.405 at

$n = 100$ and $\rho = 0.8$, M3 reaching 3.443, and M4 reaching 3.64. M1 also excels in root mean integrated squared error (RMISE), with values between 0.534 and 0.754 at $n = 100$ and $n = 400$, while M2 ranges from 0.579 to 0.782. At $n = 100$, M3 and M4 have even higher values, further demonstrating M1's effectiveness in making accurate predictions.

At $\sigma = 1.0$, similar trends are observed with the performance of M1 being even more pronounced. For instance, M1 maintains a coverage probability of 1.00 when $\rho = 0.8$. The ISPE for M1 ranges from 2.280 to 5.771, while M4's ISPE peaks at 6.681, highlighting M1's advantage in prediction accuracy. Additionally, the false positive rates for M2, M3, and M4 increase, reaching more than 40 at $n = 400$, further affirming M1's superior performance. Despite maintaining a median true positive count of 4, M1 shows lower bias in parameter estimation, with Bias(ρ) ranging from -0.107 to 0.002 compared to higher biases in the other methods.

In summary, the multi-step sparse boosting method (M1) outperforms alternative approaches in variable selection and parameter estimation, excelling in maintaining lower false positive rates and superior estimation accuracy. This establishes the multi-step sparse boosting method as a highly effective method for high-dimensional data analysis, particularly in scenarios involving complex relationships among variables.

Table 2. Results of variable screening, variable selection, and estimation for simulation when $\sigma = 1$. The values in the parentheses are the robust standard deviations. M1: our proposed multi-step sparse boosting method; M2: except sparse boosting, l_2-boosting are used in each step; M3: except sparse boosting, lasso regression are used in each step; M4: except sparse boosting, elastic net regression are used in each step. S: coverage probability that the top $\lfloor \frac{n}{\log(n)} \rfloor$ covariates after screening includes all important covariates; TP: the median of true positive; FP: the median of false positive; Size: the median of model sizes; ISPE: the average of mean square error; RMISE: the average of root mean integrated squared error; Bias(ρ): the mean bias of ρ; Bias(σ): the mean bias of σ. The values in the parentheses are the robust standard deviations; Simulation based on 500 replicates.

n	ρ	Method	S	TP	FP	Size	ISPE	RMISE	Bias(ρ)	Bias(σ)
100	0.2	M1	0.85 (0.36)	4 (0.38)	6 (2.65)	10 (2.67)	2.280 (1.853)	1.000 (0.537)	-0.079 (0.329)	0.402 (0.503)
		M2	0.85 (0.36)	4 (0.15)	14 (1.91)	17 (1.92)	2.586 (1.790)	1.037 (0.493)	-0.072 (0.309)	0.523 (0.470)
		M3	0.85 (0.36)	4 (0.17)	15 (2.14)	19 (2.20)	2.796 (1.752)	1.128 (0.489)	-0.074 (0.319)	0.598 (0.462)
		M4	0.85 (0.36)	4 (0.15)	16 (1.73)	20 (1.79)	2.948 (1.726)	1.173 (0.479)	-0.074 (0.316)	0.650 (0.452)
	0.5	M1	0.90 (0.30)	4 (0.19)	6 (2.49)	10 (2.44)	2.656 (1.762)	0.921 (0.429)	-0.004 (0.186)	0.335 (0.387)
		M2	0.90 (0.30)	4 (0.10)	14 (1.85)	18 (1.84)	3.067 (1.799)	0.980 (0.413)	0.003 (0.174)	0.471 (0.371)
		M3	0.90 (0.30)	4 (0.08)	16 (1.80)	20 (1.80)	3.323 (1.865)	1.070 (0.418)	0.003 (0.184)	0.544 (0.374)
		M4	0.90 (0.30)	4 (0.05)	16 (1.44)	20 (1.44)	3.446 (1.726)	1.116 (0.408)	0.001 (0.181)	0.594 (0.360)
	0.8	M1	1 (0.06)	4 (0)	6 (2.16)	10 (2.16)	5.771 (2.019)	0.782 (0.186)	0.002 (0.058)	0.190 (0.164)
		M2	1 (0.06)	4 (0)	13 (1.76)	17 (1.76)	6.277 (1.939)	0.824 (0.162)	0.004 (0.056)	0.313 (0.159)
		M3	1 (0.06)	4 (0)	15 (1.92)	19 (1.92)	6.567 (1.921)	0.906 (0.172)	0.008 (0.056)	0.373 (0.162)
		M4	1 (0.06)	4 (0)	16 (1.37)	20 (1.37)	6.681 (1.848)	0.943 (0.179)	0.006 (0.055)	0.424 (0.168)
400	0.2	M1	1 (0)	4 (0.65)	12 (4.03)	16 (4.38)	2.327 (1.895)	0.715 (0.386)	-0.051 (1.052)	0.440 (0.526)
		M2	1 (0)	4 (0)	42 (3.43)	46 (3.43)	2.367 (1.498)	0.715 (0.089)	-0.012 (0.939)	0.451 (0.390)
		M3	1 (0)	4 (0)	45 (12.00)	49 (12.00)	2.245 (1.363)	0.727 (0.157)	-0.005 (1.020)	0.460 (0.418)
		M4	1 (0)	4 (0)	47 (10.79)	51 (10.79)	2.304 (1.337)	0.762 (0.151)	-0.036 (1.047)	0.494 (0.432)
	0.5	M1	1 (0)	4 (0.43)	12 (3.86)	16 (4.01)	2.484 (1.819)	0.665 (0.258)	-0.080 (0.594)	0.361 (0.416)
		M2	1 (0)	4 (0)	41 (3.60)	45 (3.60)	2.635 (1.733)	0.714 (0.092)	-0.069 (0.599)	0.444 (0.411)
		M3	1 (0)	4 (0)	44 (11.66)	48 (11.66)	2.579 (1.601)	0.715 (0.154)	-0.085 (0.648)	0.432 (0.407)
		M4	1 (0)	4 (0)	46 (10.75)	50 (10.75)	2.650 (1.606)	0.751 (0.148)	-0.039 (0.613)	0.453 (0.387)
	0.8	M1	1 (0)	4 (0)	12 (3.38)	16 (3.38)	4.800 (1.832)	0.628 (0.081)	-0.107 (0.227)	0.330 (0.382)
		M2	1 (0)	4 (0)	41 (3.81)	45 (3.81)	5.042 (1.664)	0.687 (0.056)	-0.111 (0.195)	0.412 (0.347)
		M3	1 (0)	4 (0)	44 (12.22)	48 (12.22)	5.101 (1.827)	0.682 (0.120)	-0.109 (0.236)	0.415 (0.401)
		M4	1 (0)	4 (0)	45 (10.98)	49 (10.98)	5.152 (1.787)	0.716 (0.102)	-0.105 (0.229)	0.432 (0.376)

4. Real Data Analysis

In this section, we apply the proposed method to the Boston housing price data, originally collected by Harrison and Rubinfield [46] and integrated by Gilley and Pace [47]. The dataset is available for download from http://ugrad.stat.ubc.ca/R/library/mlbench/html/BostonHousing.html. It comprises median house prices observed in 506 census tracts in the Boston area in 1970, alongside a set of variables presumed to influence house prices. Table 3 provides a detailed description of these variables:

Table 3. Description of the variables in Boston housing data.

Variable	Varaible Description
MEDV	Median value of owner-occupied housing expressed in USD 1,000's
CRIM	Per capita murder rate by town
ZN	Proportion of residential land zoned for lots over $25,000$ square feet
B	Proportion of Black residents by town
RM	Average number of rooms per dwelling
DIS	Weighted distances to five Boston employment centers
NOX	Nitric oxides concentration (parts per 10 millions) per town
AGE	Proportion of owner-occupied units built before 1940
INDUS	Proportion of non-retail business acres per town
RAD	Index of accessibility to radial highways per town
PTRATIO	Pupil-teacher ratio by town
LSTAT	Percentage of lower status population
TAX	Full-value property tax rate per USD 10,000
CHAS	Charles River dummy variable (1 if tract bounds river; 0 otherwise)

The Boston housing price dataset is widely used in spatial econometrics and has been extensively studied in the literature. For the same data set, Xie et al. [48] adopted spatial autoregressive model, Li and Mei [13] considered partially linear spatial autoregressive model and Du et al. [16] used partially linear additive spatial autoregressive models. In the following, we will consider the additive spatial autoregressive model

$$
\begin{aligned}
\text{MEDV}_i =& \rho \sum_{j=1}^{506} w_{ij}\text{MEDV}_j + g_1(\text{CRIM}) + g_2(\text{ZN}) + g_3(\text{B}) + g_4(\text{RM}) + g_5(\text{DIS}) + g_6(\text{NOX}) + g_7(\text{AGE}) \\
&+ g_8(\text{INDUS}) + g_9(\text{RAD}) + g_{10}(\text{PTRATIO}) + g_{11}(\text{LSTAT}) + g_{12}(\text{TAX}) + \epsilon_i, i = 1, \cdots, 506.
\end{aligned}
\tag{18}
$$

Where MEDV_i represents the median house value for census tract i, ρ is the spatial autoregressive parameter, w_{ij} denotes the spatial weight matrix with entries set as $\max(1 - \frac{d_{ij}}{d_0}, 0)$, where d_{ij} is the Euclidean distance based on longitude and latitude coordinates of any two houses, d_0 is the threshold distance (set to 0.05 following Su and Yang [4]), resulting in a weight matrix with 19.1% nonzero elements. For comparative analysis, we consider the following methods explored in our simulation study: multi-step sparse boosting, multi-step l_2-boosting, multi-step lasso and multi-step elastic net. These methods will be used to estimate the functions $g_j(.)$ and ρ, aiming to identify the significant determinants of housing prices while accounting for spatial dependencies among census tracts.

The results of variable selection, estimation and prediction are summarized in Table 4. In addition to the number of selected variables, we also report the in-sample prediction error (ISPE) $\sum_{i=1}^n (Y_i - \hat{Y}_i)^2$ and out-of-sample prediction error (OSPE) measured by 5-fold cross-validation.

From Table 4, we observe that, compared to traditional lasso and elastic net methods, the l_2 Boosting and Sparse l_2 Boosting methods for the varying-coefficient model exhibit smaller OSPE, while also producing relatively sparser models. The smaller ISPE may be due to overfitting in the traditional models which choosing all of the variables. These conventional models exhibit errors that increase by a factor of 6 to 7 during out-of-sample predictions, whereas our proposed model demonstrates greater stability with smaller fluctuations when comparing ISPE to OSPE. Our Sparse L_2 Boosting method for the varying-coefficient AFT model performs remarkably well in terms of sparsity, estimation, and prediction.

Table 4. Boston housing price example: variable selection, estimation and prediction performance. No.: Number of variables selected; Variables: Names of selected variables; ISPE: In-sample prediction error $\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$; OSPE: 5-fold cross-validation out-of-sample prediction error.

Method	No.	Variables	ISPE	OSPE
multi-step sparse boosting	2	RM (3), LSTAT (9)	0.665	0.951
multi-step l_2 boosting	2	RM (3), LSTAT (9)	0.665	0.951
multi-step lasso	12	CRIM (1), B (2), RM (3), DIS (4), NOX (5), AGE (6), INDUS (7), PTRATIO (8), LSTAT (9), ZN (10), RAD (11), TAX (12)	0.172	1.197
multi-step elastic net	12	CRIM (1), B (2), RM (3), DIS (4), NOX (5), AGE (6), INDUS (7), PTRATIO (8), LSTAT (9), ZN (10), RAD (11), TAX (12)	0.160	1.232

Additionally, we present a Venn diagram in Figure 2 showing the overlapping genes identified by all different methods. We observe that SparseL_2Boosting and l_2Boosting select 2 variables (RM and LSTAT) which mostly affect the Boston housing prices in common, while lasso and elastic net methods identify all the 12 variables in common. However, SparseL_2Boosting and l_2Boosting produce distinct lists of selected variables compared to the other methods, with only 2 variables selected by all methods.

We plot the estimated curves of varying coefficients for RM and LSTAT, which are the two selected variables by sparse boosting method, with their 95% confidence bands constructed by 500 bootstrap resamples in Figure 3. All of the functions are quite different from a straight line and this suggest that the varying-coefficient model is more appropriate to describe the covariate effects on boston housing prices in our data. Using more sophisticated semiparametric model specification may provide more accurate model for the high-dimensional analysis. Furthermore, we observe different functional forms for the two impact variables.
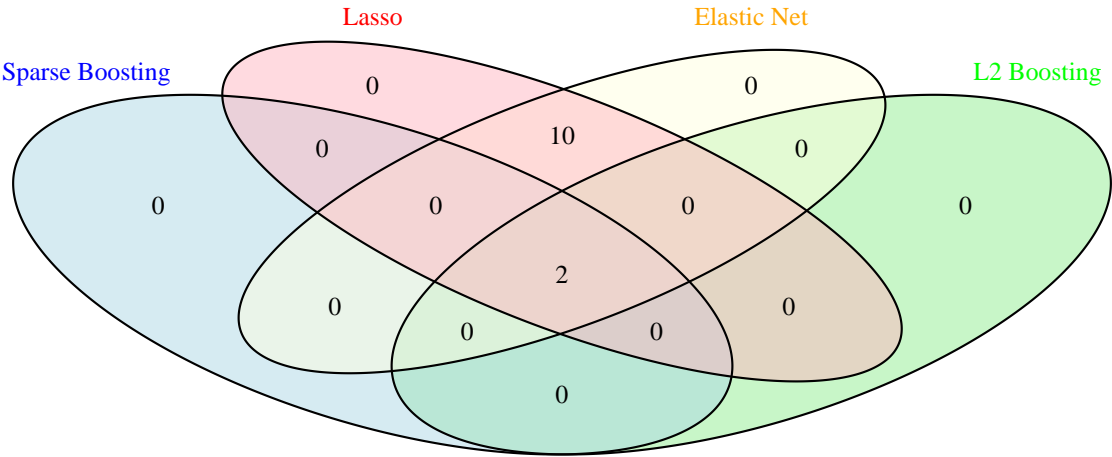

Figure 2. Venn diagram showing overlapping genes selected among all methods. Blue: sparse boosting; Green: l_2 boosting; Red:lasso regression; Yellow: elastic net regression
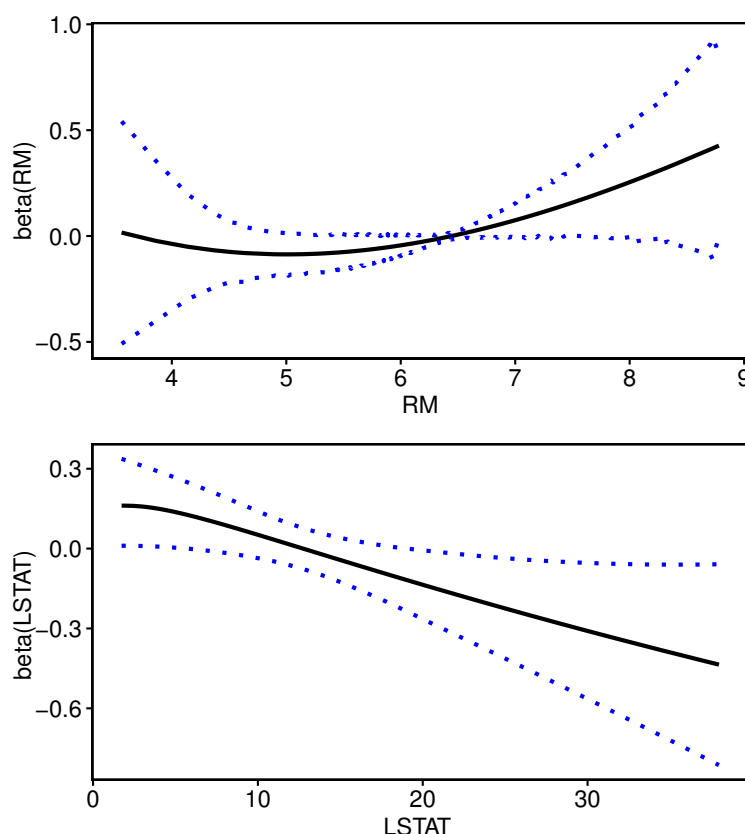
Figure 3. Estimated nonparametric function with their 95% confidence bands based on 500 bootstraps for the selected variables in the final spatial additive model by multi-step sparse boosting.

5. Concluding Remarks

This paper introduces a useful multi-step sparse boosting algorithm specifically designed for additive spatial autoregressive models with high-dimensional covariates, addressing critical aspects of variable selection and parameter estimation. Our methodology effectively facilitates model-based prediction without the burden of time-consuming tuning parameter selection, thereby streamlining the modeling process.

Using B-spline basis functions for approximating varying coefficients is a critical component of our approach, which allows for smooth and accurate representations even when the underlying assumptions regarding smoothness may be violated, enhancing the robustness of the model. Moreover, a noteworthy aspect of our finding is the two-step sparse boosting approach employing the generalized Minimum Description Length (gMDL) model selection criterion. The gMDL criterion utilizes a data-driven penalty for squared loss, promoting model parsimony while effectively filtering out irrelevant covariates.

The simulation results illustrate the efficacy of our proposed approach. Our multi-step sparse boosting method consistently exhibits fewer false positives than alternative methods, while true positive rates accurately reflect the actual number of relevant variables. In contrast, traditional methods utilizing boosting, lasso, or elastic net tend to suffer from overfitting, as evidenced by lower in-sample prediction errors (ISPE) and larger model sizes. Meanwhile, the real-world application of our methodology to the Boston housing price data further confirms its effectiveness and stability in identifying significant predictors of housing prices. By accurately selecting key variables, our approach enhances the model's predictive performance and interpretability.

However, several limitations of our study must be acknowledged. Firstly, although B-spline basis functions offer significant advantages, they may not fully capture the intricate complexities of all data relationships, potentially limiting the model's applicability to more diverse or nuanced datasets.

Therefore, while our research encompasses thorough simulations and empirical testing on the housing price datasets, it is imperative to extend evaluations across a wider array of datasets.

This method can be extended to various fields, not limited to environmental studies, healthcare, and social sciences. Additionally, it is applicable to both nonparametric and semiparametric models, although our focus here has been on nonparametric approaches. This broader analysis is crucial to comprehensively validate the model's versatility and effectiveness across various contexts within the disciplines of econometrics and statistics.

References

1. C. Andrew, J. K. Ord, Spatial autocorrelation, Pion, London (1973).
2. L. Anselin, Spatial econometrics: methods and models (vol. 4), Studies in Operational Regional Science. Dordrecht: Springer Netherlands (1988). http://dx.doi.org/10.1007/978-94-015-7799-1_2
3. R. Haining, Spatial data analysis in the social and environmental sciences, Cambridge University Press, 1993. http://dx.doi.org/10.1017/cbo9780511623356
4. L. Su, Z. Yang, Instrumental variable quantile estimation of spatial autoregressive models (2007). http://econpapers.repec.org/paper/eabdevelo/1563.htm
5. H. H. Kelejian, I. R. Prucha, Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, Journal of econometrics 157 (1) (2010) 53–67. http://dx.doi.org/10.1016/j.jeconom.2009.10.025
6. X. Lin, L.-f. Lee, Gmm estimation of spatial autoregressive models with unknown heteroskedasticity, Journal of Econometrics 157 (1) (2010) 34–52. http://dx.doi.org/10.1016/j.jeconom.2009.10.035
7. X. Liu, L.-f. Lee, C. R. Bollinger, An efficient gmm estimator of spatial autoregressive models, Journal of Econometrics 159 (2) (2010) 303–319. http://dx.doi.org/10.1016/j.jeconom.2010.08.001
8. H. Li, C. A. Calder, N. Cressie, One-step estimation of spatial dependence parameters: Properties and extensions of the aple statistic, Journal of Multivariate Analysis 105 (1) (2012) 68–84. http://dx.doi.org/10.1016/j.jmva.2011.08.006
9. L. Su, S. Jin, Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models, Journal of Econometrics 157 (1) (2010) 18–33. http://dx.doi.org/10.1016/j.jeconom.2009.10.033
10. M. Koch, T. Krisztin, Applications for asynchronous multi-agent teams in nonlinear applied spatial econometrics, Journal of Internet Technology 12 (6) (2011) 1007–1014. https://doi.org/10.6138/JIT.2011.12.6.14
11. W. R. Chen, Jiaqing, H. Yangxin, Semiparametric spatial autoregressive model : A two-step bayesian approach, Annals of Public Health Research 2 (1) (2015) 1012. https://doi.org/10.47739/2378-9328/1012
12. T. Krisztin, The determinants of regional freight transport: a spatial, semiparametric approach, Geographical Analysis 49 (3) (2017) 268–308. http://dx.doi.org/10.1111/gean.12125
13. T. Li, C. Mei, Statistical inference on the parametric component in partially linear spatial autoregressive models, Communications in Statistics-Simulation and Computation 45 (6) (2016) 1991–2006. http://dx.doi.org/10.1080/03610918.2014.889152
14. L. Su, Semiparametric gmm estimation of spatial autoregressive models, Journal of Econometrics 167 (2) (2012) 543–560. http://dx.doi.org/10.1016/j.jeconom.2011.09.034
15. H. Wei, Y. Sun, M. Hu, Model selection in spatial autoregressive models with varying coefficients, Frontiers of Economics in China 13 (4) (2018) 559–576. https://doi.org/10.3868/s060-007-018-0026-2
16. J. Du, X. Sun, R. Cao, Z. Zhang, Statistical inference for partially linear additive spatial autoregressive models, Spatial statistics 25 (2018) 52–67. http://dx.doi.org/10.1016/j.spasta.2018.04.008
17. S. Cheng, J. Chen, X. Liu, Gmm estimation of partially linear single-index spatial autoregressive model, Spatial Statistics (2019). http://dx.doi.org/10.1016/j.spasta.2019.04.002
18. R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society: Series B (Methodological) 58 (1) (1996) 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x
19. J. Fan, R. Li, Variable selection via nonconcave penalized likelihood and its oracle properties, Journal of the American statistical Association 96 (456) (2001) 1348–1360. http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x

20. C.-H. Zhang, et al., Nearly unbiased variable selection under minimax concave penalty, The Annals of statistics 38 (2) (2010) 894–942. http://dx.doi.org/10.1214/09-aos729

21. H. Zou, T. Hastie, Regularization and variable selection via the elastic net, Journal of the royal statistical society: series B (statistical methodology) 67 (2) (2005) 301–320. http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x

22. H. Zou, The adaptive lasso and its oracle properties, Journal of the American statistical association 101 (476) (2006) 1418–1429. https://doi.org/10.1198/016214506000000735

23. M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (1) (2006) 49–67. http://dx.doi.org/10.1111/j.1467-9868.2005.00532.x

24. R. E. Schapire, The strength of weak learnability, Machine learning 5 (2) (1990) 197–227. http://dx.doi.org/10.1007/bf00116037

25. Y. Freund, Boosting a weak learning algorithm by majority, Information and computation 121 (2) (1995) 256–285. http://dx.doi.org/10.1006/inco.1995.1136

26. Y. Freund, R. E. Schapire, A decision-theoretic generalization of on-line learning and an application to boosting, Journal of computer and system sciences 55 (1) (1997) 119–139. http://dx.doi.org/10.1006/jcss.1997.1504

27. P. Bühlmann, B. Yu, Boosting with the l 2 loss: regression and classification, Journal of the American Statistical Association 98 (462) (2003) 324–339. https://doi.org/10.1198/016214503000125

28. P. Bühlmann, B. Yu, Sparse boosting, Journal of Machine Learning Research 7 (Jun) (2006) 1001–1024. http://jmlr.org/papers/v7/buehlmann06a.html

29. Z. Wang, Hingeboost: Roc-based boost for classification and variable selection, The International Journal of Biostatistics 7 (1) (2011) 1–30. http://dx.doi.org/10.2202/1557-4679.1304

30. O. Komori, S. Eguchi, A boosting method for maximizing the partial area under the roc curve, BMC bioinformatics 11 (1) (2010) 314. http://dx.doi.org/10.1186/1471-2105-11-314

31. P. Bühlmann, T. Hothorn, Twin boosting: improved feature selection and prediction, Statistics and Computing 20 (2) (2010) 119–138. http://dx.doi.org/10.1007/s11222-009-9148-5

32. Y. Yang, H. Zou, Nonparametric multiple expectile regression via er-boost, Journal of Statistical Computation and Simulation 85 (7) (2015) 1442–1458. https://doi.org/10.1080/00949655.2013.876024

33. J. Zhao, General sparse boosting: Improving feature selection of l2 boosting by correlation-based penalty family, Communications in Statistics-Simulation and Computation 44 (6) (2015) 1612–1640. http://dx.doi.org/10.1080/03610918.2013.824586

34. M. Yue, J. Li, S. Ma, Sparse boosting for high-dimensional survival data with varying coefficients, Statistics in medicine 37 (5) (2018) 789–800. http://dx.doi.org/10.1002/sim.7544

35. M. Yue, J. Li, M.-Y. Cheng, Two-step sparse boosting for high-dimensional longitudinal data with varying coefficients, Computational Statistics & Data Analysis 131 (2019) 222–234. http://dx.doi.org/10.1016/j.csda.2018.10.002

36. M. Yue, L. Huang, A new approach of subgroup identification for high-dimensional longitudinal data, Journal of Statistical Computation and Simulation 90 (11) (2020) 2098–2116. http://dx.doi.org/10.1080/00949655.2020.1764555

37. M. Yue, J. Li, B. Sun, Conditional sparse boosting for high-dimensional instrumental variable estimation, Journal of Statistical Computation and Simulation 92 (15) (2022) 3087–3108. http://dx.doi.org/10.1080/00949655.2022.2056739

38. J. Huang, J. L. Horowitz, F. Wei, Variable selection in nonparametric additive models, Annals of statistics 38 (4) (2010) 2282. http://dx.doi.org/10.1214/09-aos781

39. Y. Zhang, D. Shen, Estimation of semi-parametric varying-coefficient spatial panel data models with random-effects, Journal of Statistical Planning and Inference 159 (2015) 64–80. http://dx.doi.org/10.1016/j.jspi.2014.11.001

40. M. H. Hansen, B. Yu, Model selection and the principle of minimum description length, Journal of the American Statistical Association 96 (454) (2001) 746–774. https://doi.org/10.1198/016214501753168398

41. L.-F. Lee, Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models, Econometrica 72 (6) (2004) 1899–1925. http://dx.doi.org/10.1111/j.1468-0262.2004.00558.x

42. M. Yue, J. Li, Improvement screening for ultra-high dimensional data with censored survival outcomes and varying coefficients, The international journal of biostatistics 13 (1) (2017). http://dx.doi.org/10.1515/ijb-2017-0024

43. J. Fan, J. Lv, Sure independence screening for ultrahigh dimensional feature space, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 70 (5) (2008) 849–911. http://dx.doi.org/10.1111/j.1467-9868.2008.00674.x

44. M.-Y. Cheng, T. Honda, J. Li, H. Peng, et al., Nonparametric independence screening and structure identification for ultra-high dimensional longitudinal data, The Annals of Statistics 42 (5) (2014) 1819–1849. http://dx.doi.org/10.1214/14-aos1236

45. X. Xia, J. Li, B. Fu, Conditional quantile correlation learning for ultrahigh dimensional varying coefficient models and its application in survival analysis, Statistica Sinica (2018). http://dx.doi.org/10.5705/ss.202016.0402

46. D. Harrison Jr, D. L. Rubinfeld, Hedonic housing prices and the demand for clean air, Journal of environmental economics and management 5 (1) (1978) 81–102. http://dx.doi.org/10.1016/0095-0696(78)90006-2

47. O. W. Gilley, R. K. Pace, et al., On the harrison and rubinfeld data, Journal of Environmental Economics and Management 31 (3) (1996) 403–405. http://dx.doi.org/10.1006/jeem.1996.0052

48. T. Xie, R. Cao, J. Du, Variable selection for spatial autoregressive models with a diverging number of parameters, Statistical Papers (2018) 1–21. http://dx.doi.org/10.1007/s00362-018-0984-2